

Aplicações Informáticas na Biomedicina

5.^a Aula Prática Laboratorial

Mestrado Integrado em Engenharia Informática

Ano Letivo 2019/2020

Marisa Esteves

23 de Outubro de 2019



Universidade do Minho

Plano de Aula

1. Finalização da resolução da 3.^a ficha prática laboratorial pelos alunos em grupo;
2. Correção da ficha com os alunos;
3. Definição dos grupos de trabalho;
4. Instalação do Talend.

Processo ETL

Definição

O processo ETL (*Extract, Transform, Load*) é um conjunto de processos que inclui a extração de dados de fontes de informação internas e externas, podendo estar em diferentes formatos, a transformação dos dados de acordo com as necessidades da organização e, finalmente, o carregamento dos mesmos numa estrutura de dados, como por exemplo um data mart ou um data warehouse.

Processo ETL

Definição

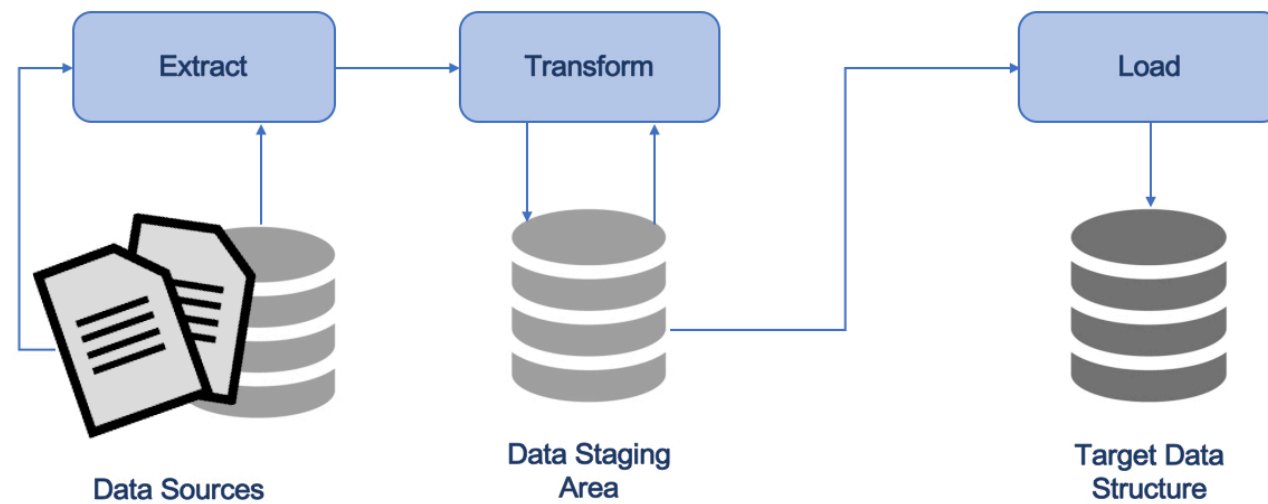


Figura 1 – Esquema do processo ETL.

Processo ETL

Porquê?

Os dados estão espalhados
por diferentes localizações

Os dados estão
armazenados em diferentes
tipos de formato

O volume de dados continua
a aumentar

Os dados podem estar
estruturados, semi-
estruturados ou não
estruturados

Data Warehousing

Definição

O processo de data warehousing enfatiza à recolha de dados de diversas fontes através do processo ETL (*Extract, Transform, Load*), correspondendo à construção de data warehouses e/ou data marts, para aceder e analisar a informação de forma útil. Os dados extraídos são processados, formatados e consolidados numa estrutura de dados única para facilitar essencialmente a análise de dados.

Data Warehousing

Definição

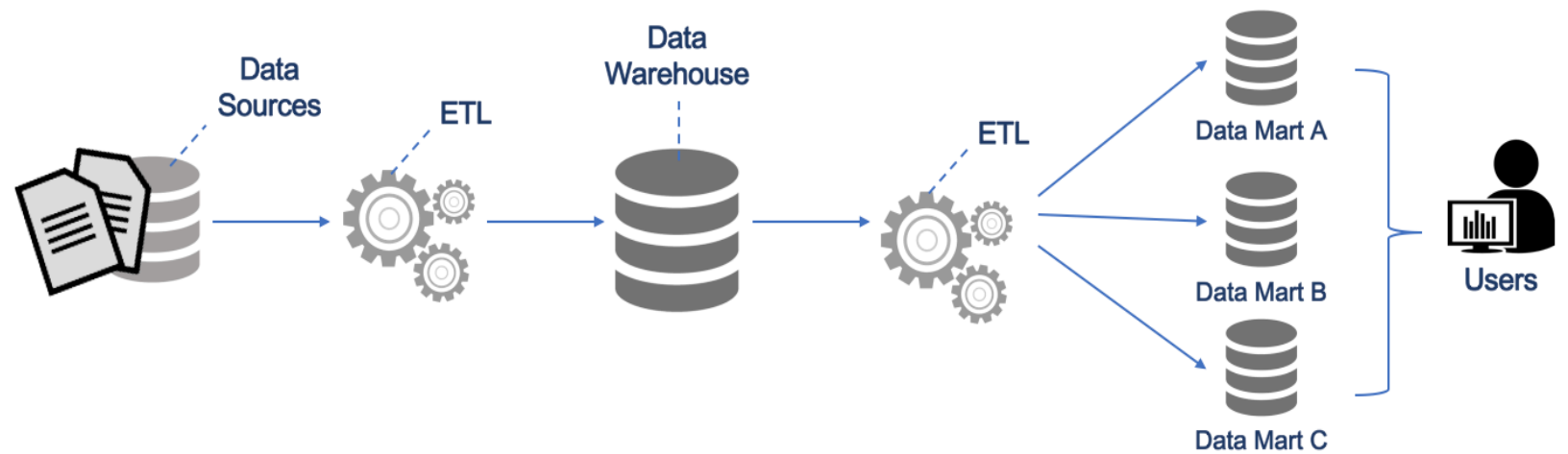


Figura 2 – Esquema do processo de data warehousing.

Data Warehousing

*Data Warehouse vs. Data
Mart*

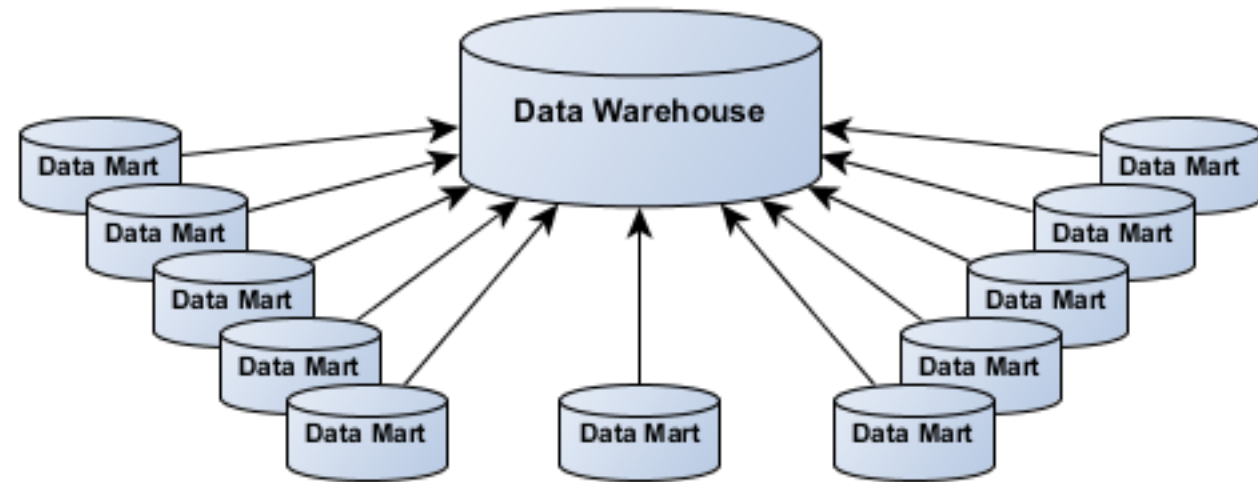


Figura 3 – Data warehouse vs. Data marts.

Data Warehousing

*Modelo Dimensional –
Esquema em Estrela vs.
Esquema em Floco de Neve*

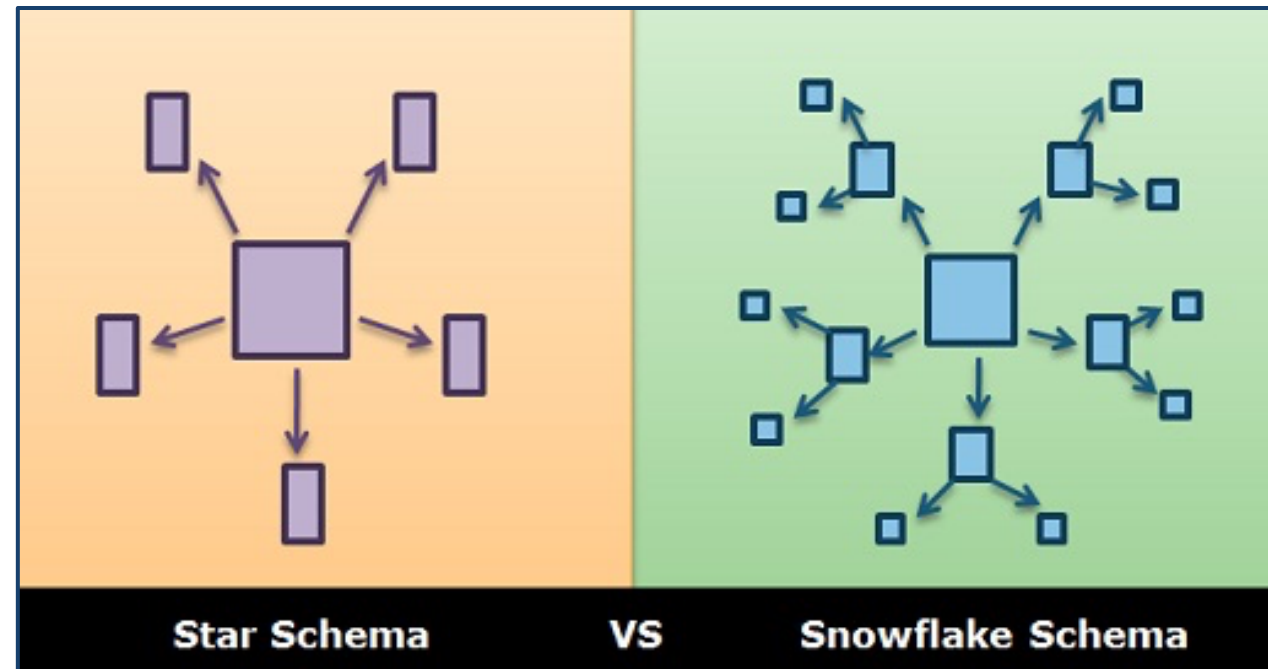


Figura 4 – Esquema em Estrela vs. Esquema em Floco de Neve.

Data Warehousing

*Modelo Dimensional –
Esquema em Constelação
de Factos*

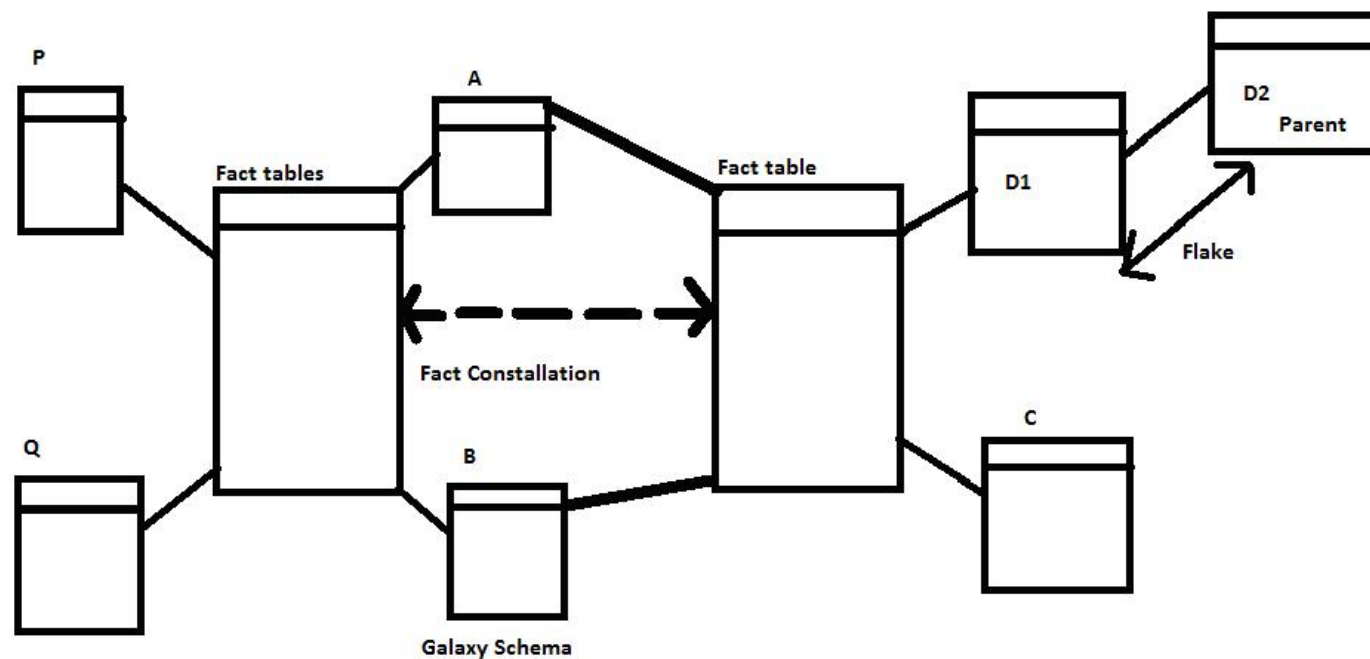


Figura 5 – Esquema em Constelação de Factos.

OLTP vs. OLAP

Definição

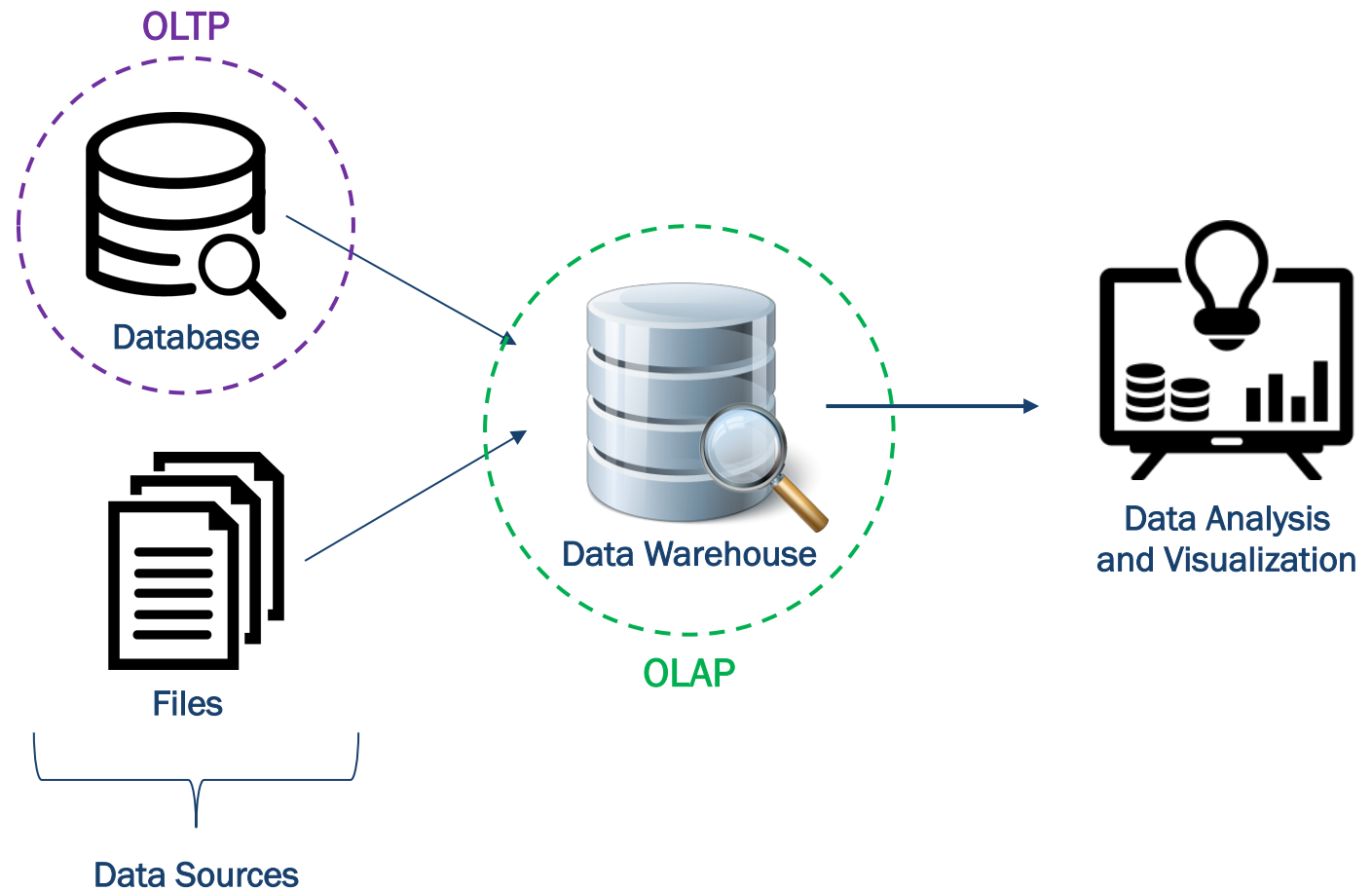


Figura 6 – OLTP (*Online Transaction Processing*) vs. OLAP (*Online Analytical Processing*).

OLTP vs. OLAP

Definição

Relational Database (OLTP)	Analytical Data Warehouse (OLAP)
Contains current data	Contains historical data
Useful in running the business	Useful in analysing the business
Based on Entity Relationship Model	Based on Star, Snowflake or Fact Constellation Schema
Provides primitive and highly detailed data	Provides summarized and consolidated data
Used for writing into the database	Used for reading data from the data warehouse
Database size ranges from 100 MB to 1 GB	Data warehouse ranges from 100 GB to 1 TB
Fast and it provides high performance	Highly flexible but it is not fast
Number of records accessed is in tens	Number of records accessed is in millions
Example: all bank transactions made by a customer	Example: bank transactions made by a customer at a particular time

Figura 7 – Diferenças entre OLTP e OLAP.

MySQL

**INSERT INTO
SELECT FROM**

*Permite copiar dados de uma tabela e os inserir noutra tabela.
No entanto, este comando SQL requer que os tipos de dados na tabela
de origem (table1) e na tabela destino (table2) sejam iguais.*

- **INSERT INTO** *table2* (*column1*, *column2*, *column3*, ...) **SELECT** *column1*, *column2*, *column3*, ...
FROM *table1*
WHERE *condition*

MySQL

Cursores

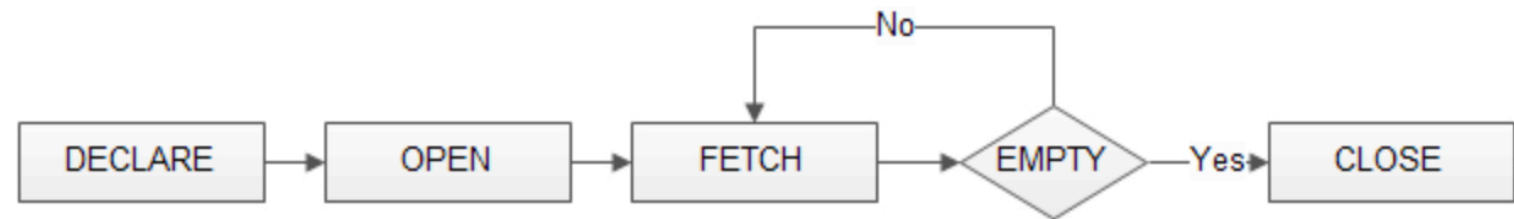


Figura 8 – Modo de funcionamento de cursores em MySQL.

MySQL

Cursores

```
1 CREATE PROCEDURE curdemo()  
2 BEGIN  
3     DECLARE done INT DEFAULT FALSE;  
4     DECLARE a CHAR(16);  
5     DECLARE b, c INT;  
6     DECLARE cur1 CURSOR FOR SELECT id,data FROM test.t1;  
7     DECLARE cur2 CURSOR FOR SELECT i FROM test.t2;  
8     DECLARE CONTINUE HANDLER FOR NOT FOUND SET done = TRUE;  
9  
10    OPEN cur1;  
11    OPEN cur2;  
12  
13    read_loop: LOOP  
14        FETCH cur1 INTO a, b;  
15        FETCH cur2 INTO c;  
16        IF done THEN  
17            LEAVE read_loop;  
18        END IF;  
19        IF b < c THEN  
20            INSERT INTO test.t3 VALUES (a,b);  
21        ELSE  
22            INSERT INTO test.t3 VALUES (a,c);  
23        END IF;  
24    END LOOP;  
25  
26    CLOSE cur1;  
27    CLOSE cur2;  
28 END;
```

Figura 9 – Exemplo de um procedimento com cursores em MySQL.

Resolução da 3.ª Ficha Prática Laboratorial

1 Modelo Dimensional de uma Lista de Espera de Cirurgias

O ficheiro disponibilizado juntamente com esta ficha prática laboratorial, nomeadamente `lista_espera_blo.csv`, contém dados reais da lista de espera de cirurgias num determinado hospital nacional. A informação é guardada numa base de dados nos sistemas de informação hospitalar da instituição de saúde. Os dados de cerca de 1000 registos foram extraídos de uma tabela da base de dados para um ficheiro no formato `.csv`.

A informação representada inclui o número na lista de espera (identificador único), o número sequencial do paciente, o número do processo, a data de marcação, a data de cancelamento, a data de realização da cirurgia, a descrição do grupo, a descrição da patologia e a descrição da intervenção cirúrgica.

É importante referir que, por razões de confidencialidade, determinada informação do paciente foi retirada do *dataset* inicial (colunas).

Os profissionais de tecnologias de informação do hospital pretendem remodelar a organização da informação em questão num modelo dimensional baseado no esquema em estrela.

Com base no caso apresentado, pretende-se que:

1. Crie um novo *schema* no MySQL Workbench denominado “Ficha3”.
2. Faça o *import* dos dados no ficheiro `lista_espera_blo.csv` para uma nova tabela no *schema* criado na alínea anterior. No processo, uma tabela denominada “`lista_espera_blo`” deverá ser criada e povoada corretamente com os dados do ficheiro (Table Data Import Wizard).

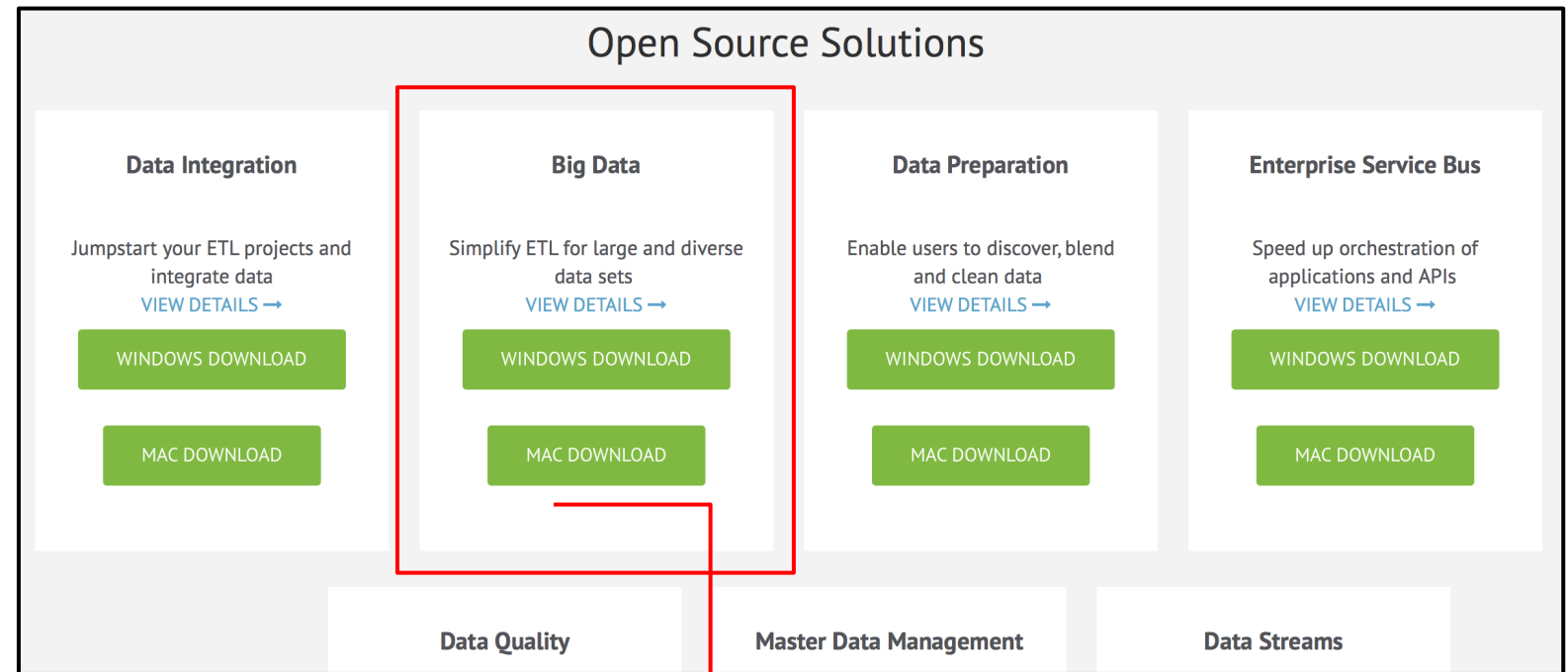
Resolução da 3.ª Ficha Prática Laboratorial

3. Analise a estrutura da tabela `lista_espera_blo` e, conseqüentemente, defina um modelo dimensional no formato de esquema em estrela. O modelo deverá ter uma tabela de factos e as respetivas tabelas de dimensão ligadas à tabela de factos.
4. Construa o modelo dimensional definido na alínea anterior no MySQL Workbench (*EER diagram*).
5. Faça a conversão do modelo lógico criado para o respetivo modelo físico para o *schema* Ficha3 (Database > Forward Engineer).
6. Povoie todas as tabelas do modelo dimensional (tabela de factos e tabelas de dimensão) em SQL a partir da tabela `lista_espera_blo`. É de lembrar que as tabelas de dimensão deverão ser povoadas antes da tabela de factos.

Instalação

Talend

<https://www.talend.com/products/talend-open-studio/>



+ third party libraries (jars)